

IFSP - Instituto Federal de Educação, Ciência e Tecnologia
Câmpus São Paulo

Kevin Klein SP3096289

Ruan de Souza SP3093697

Pedro Dias SP3099211

Projeto - Análise Exploratória de Dados

Professor: Dra. Josceli M. Tenorio

IFSP - Instituto Federal de Educação, Ciência e Tecnologia
Câmpus São Paulo

Tecnologia em Análise e Desenvolvimento de Sistemas

ESP1A5 - Estatística e Probabilidade

São Paulo - SP - Brasil

2024

Sumário

1	INTRODUÇÃO	3
1.0.1	Descrição das Bases de Dados	3
2	ANÁLISE	5
2.0.1	Análise Descritiva	5
2.0.2	Metodologia	5
2.0.3	Variáveis Utilizadas	6
2.0.4	Análise das Hospitalizações	8
2.0.5	Análise das Vacinações	11
3	INFERÊNCIA	15
3.1	Vacinação	15
3.2	Hospitalizações	15
3.2.1	Teste de Normalidade	15
3.2.2	Comparação entre 2022 e 2023	16
3.2.3	Correlação entre Ocupação de UTI e Data	16
3.2.4	Modelo de Regressão	16
4	PROBABILIDADE	18
4.0.1	Estudo de Probabilidade	18
4.0.2	Dados sobre a vacinação	18
5	CONCLUSÃO	20
5.0.1	Desvendando a Relação entre Vacinação e Hospitalizações	20
5.0.2	Um Panorama de Hospitalizações	20
5.0.3	Vacinação	20
5.0.4	Vacinação e Redução de Hospitalizações	20
5.0.5	Reconhecendo Limitações	21
6	REFERÊNCIA BIBLIOGRÁFICA	22

1 Introdução

Este trabalho apresenta uma análise exploratória abrangente dos dados de vacinações e hospitalizações relacionadas à COVID-19, abrangendo o período desde o início de 2020 até os dias atuais. Todos os dados foram coletados do repositório covid-19-data, criado pela Our World in Data (OWID), uma publicação científica online que se dedica a abordar grandes problemas globais, como pobreza, doenças, fome, mudanças climáticas, guerra, riscos existenciais e desigualdades.

O OWID é um projeto do Global Change Data Lab, uma organização beneficente registrada na Inglaterra e no País de Gales, fundado por Max Roser, historiador social e economista do desenvolvimento. A equipe de pesquisa está sediada na Universidade de Oxford e é presidida por Hetan Shah.

Neste relatório, nosso objetivo é oferecer insights fundamentados sobre o progresso da vacinação contra a COVID-19 e a situação das hospitalizações em diferentes países, continentes e classes sociais. Essa análise visa fornecer uma compreensão mais clara do impacto das medidas de saúde pública e da eficácia das campanhas de vacinação, utilizando dados confiáveis e abrangentes disponíveis no repositório do OWID.

Vamos explorar a distribuição das vacinações em todo o mundo, identificando os países com maior número de doses administradas, bem como investigar a ocupação de leitos de UTI em diferentes regiões, enfocando as médias de ocupação ao longo do período analisado.

1.0.1 Descrição das Bases de Dados

Da vasta gama de dados estatísticos disponibilizados pelo repositório da Our World in Data sobre a COVID-19, selecionamos especificamente as bases de dados referentes à vacinação e às hospitalizações em todo o mundo. Esses conjuntos de dados abrangem informações desde o início de 2020 até os dias atuais, sendo continuamente atualizados pela organização para refletir os desenvolvimentos mais recentes.

A base de dados de hospitalizações contém cerca de 230 mil linhas e 5 colunas, agrupando informações por indicadores como Ocupação diária de UTIs e Admissões semanais dos hospitais, entre outros. Esses dados abrangem mais de 50 países, fornecendo uma visão detalhada das tendências de hospitalização ao longo do tempo e permitindo uma análise aprofundada das respostas hospitalares à pandemia.

A base de dados de vacinações, por sua vez, apresenta um nível ainda maior de detalhamento, com 17 colunas e mais de 190 mil linhas. Esta base inclui informações

de mais de 50 países, bem como dados agregados por continentes, classes sociais e dados globais. Entre as colunas, encontramos informações sobre o total de vacinações, vacinações diárias, número de pessoas que receberam duas doses, entre outros dados cruciais. Esse detalhamento permite uma análise abrangente da distribuição e eficácia das campanhas de vacinação ao redor do mundo.

No total, os dados específicos da nossa análise possuem aproximadamente 27MB, evidenciando a riqueza e a amplitude das informações disponíveis para estudo. float

2 Análise

2.0.1 Análise Descritiva

Nesta seção, realizaremos uma análise descritiva abrangente dos dados de internações hospitalares relacionadas à COVID-19 e do progresso das vacinações em diferentes regiões do mundo. Nosso objetivo é fornecer uma visão detalhada e comparativa do número de internações e vacinações, permitindo uma compreensão mais profunda da situação da pandemia e do esforço de vacinação global.

Inicialmente, exploraremos os dados de hospitalizações, examinando a contagem de internações em diferentes países e continentes. Investigaremos as tendências ao longo do tempo, identificando os períodos de aumento ou redução nas hospitalizações e destacando as regiões mais afetadas.

Em seguida, nos concentraremos nas vacinações contra a COVID-19, analisando o número total de doses administradas em várias partes do mundo. Vamos comparar diferentes países e continentes para entender como o processo de vacinação está progredindo e se há disparidades significativas entre as regiões.

2.0.2 Metodologia

Para realizar as análises e gerar os gráficos apresentados neste relatório, utilizamos o software estatístico R, uma ferramenta de análise de dados amplamente utilizada em estatística e ciência de dados. A seguir, descrevemos algumas das etapas específicas do nosso processo metodológico:

Coleta de Dados: Os dados foram obtidos do repositório da Our World in Data, que fornece uma vasta gama de informações sobre a COVID-19. Seleccionamos as bases de dados relacionadas à vacinação e hospitalizações, abrangendo um período desde o início de 2020 até os dias atuais. Essas bases de dados foram importadas para o R utilizando a função `read_delim` do pacote `readr`.

Pré-processamento dos Dados: Após a importação, realizamos o pré-processamento dos dados para garantir sua qualidade e adequação para análise. Isso incluiu a filtragem de dados irrelevantes, tratamento de valores ausentes e formatação adequada das colunas de data. Utilizamos as funções `filter` e `mutate` do pacote `dplyr` para essas tarefas.

Geração de Gráficos: Os gráficos foram desenvolvidos utilizando o pacote `ggplot2`, uma das bibliotecas mais renomadas do R para visualização de dados. Utilizamos diversas funções do `ggplot2` para criar diferentes tipos de gráficos. A função `geom_line` foi

empregada para a criação de gráficos de linha, permitindo a visualização clara de tendências ao longo do tempo. Para a representação de comparações categóricas, utilizamos a função `geom_bar`, que gera gráficos de barras eficazes. Além dessas, outras funcionalidades do `ggplot2` foram exploradas para adaptar as visualizações às necessidades específicas da análise, garantindo uma apresentação precisa e informativa dos dados.

Formatação e Apresentação: Por fim, aplicamos formatações apropriadas aos gráficos para melhorar a legibilidade e a clareza das informações. Utilizamos funções como `scale_y_continuous` com `labels = comma` do pacote `scales` para formatar os números em uma escala legível. Além disso, ajustamos os textos dos eixos e títulos dos gráficos utilizando `labs` e `theme`.

2.0.3 Variáveis Utilizadas

Base de Dados de Vacinação

As principais colunas utilizadas da base de dados de vacinação incluem:

- **location:**
Representa o país, continente, ou região específica (como o mundo) a que os dados se referem. Esta variável foi crucial para agrupar e filtrar os dados conforme necessário para diferentes visualizações.
- **date:**
Indica a data específica em que os dados foram registrados. Essa coluna foi essencial para a criação de gráficos temporais, permitindo o acompanhamento da evolução das vacinações ao longo do tempo.
- **total_vaccinations:**
Mostra o número total de doses de vacinas administradas até a data especificada. Utilizamos essa variável para identificar os locais com maior número de vacinações acumuladas.
- **people_vaccinated:**
Indica o número de pessoas que receberam pelo menos uma dose da vacina. Esta coluna foi utilizada para analisar a cobertura vacinal em diferentes regiões.
- **daily_vaccinations:**
Representa o número de doses administradas diariamente. Foi utilizada para calcular a média diária de vacinações e para observar picos de campanhas de vacinação.

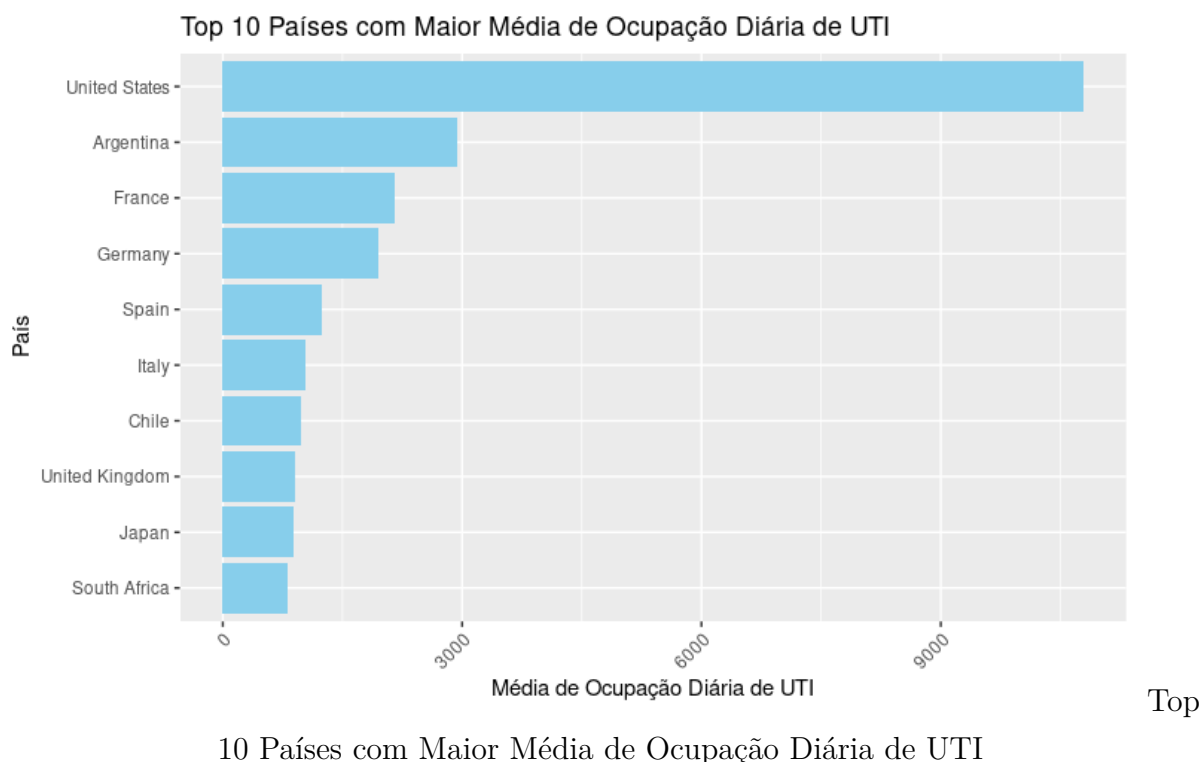
Base de Dados de Hospitalizações

As principais colunas utilizadas na base de dados de hospitalizações incluem:

- **entity:**
Similar à coluna "location" da base de vacinação, esta variável representa o país ou região de onde os dados foram coletados. Foi essencial para agrupar os dados por região.
- **date:**
Indica a data específica dos registros de hospitalização. Esta coluna permitiu a criação de gráficos que mostram a evolução temporal das hospitalizações.
- **indicator:**
Descreve o tipo de indicador de hospitalização, como "Daily ICU occupancy" (Ocupação diária de UTIs) e "Weekly new hospital admissions" (Admissões semanais nos hospitais). Utilizamos esta variável para filtrar os dados conforme o tipo de análise necessária.
- **value:**
Fornece os valores numéricos associados aos indicadores específicos. Esta variável foi utilizada diretamente nas análises quantitativas e na criação de gráficos.
- **iso_code:**
Código ISO do país, utilizado para identificar os países de forma padronizada.

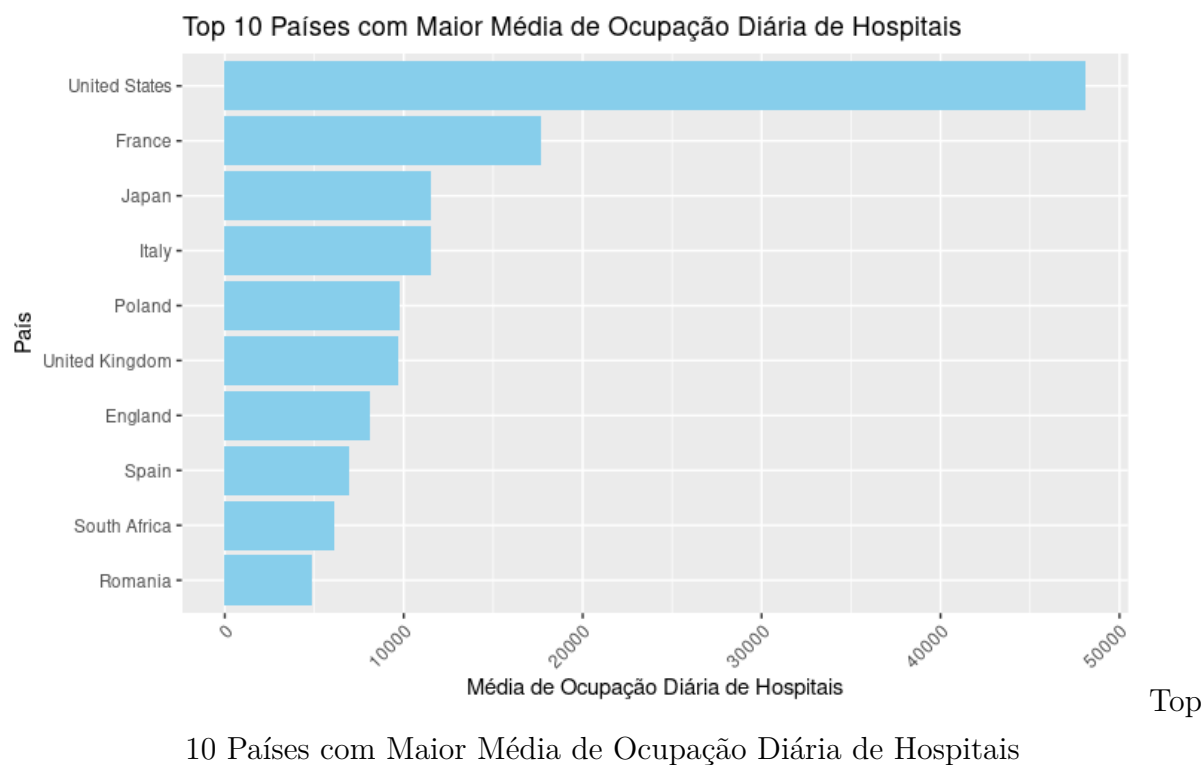
2.0.4 Análise das Hospitalizações

A Figura 1 apresenta o gráfico dos top 10 países com a maior média diária de internações na UTI, durante o período de janeiro de 2020 até janeiro de 2023, considerado o pico da pandemia. Este gráfico oferece insights sobre os países mais impactados em termos de ocupação de leitos de UTI ao longo desse período.



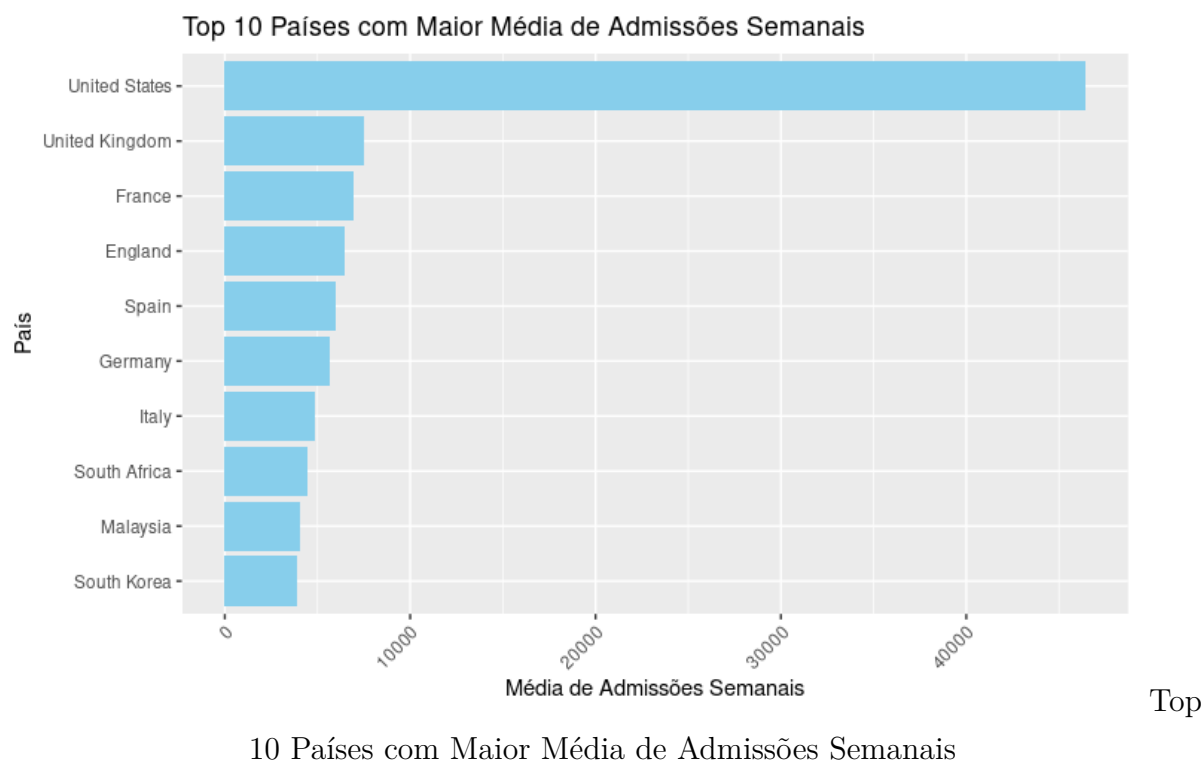
Fonte: Os autores.

A Figura 2 mostra outro gráfico semelhante, mas agora representando os top 10 países com a maior média diária total de ocupação dos hospitais durante o mesmo período. Esta visualização nos permite comparar a média total de ocupação hospitalar entre os países mais afetados pela pandemia.



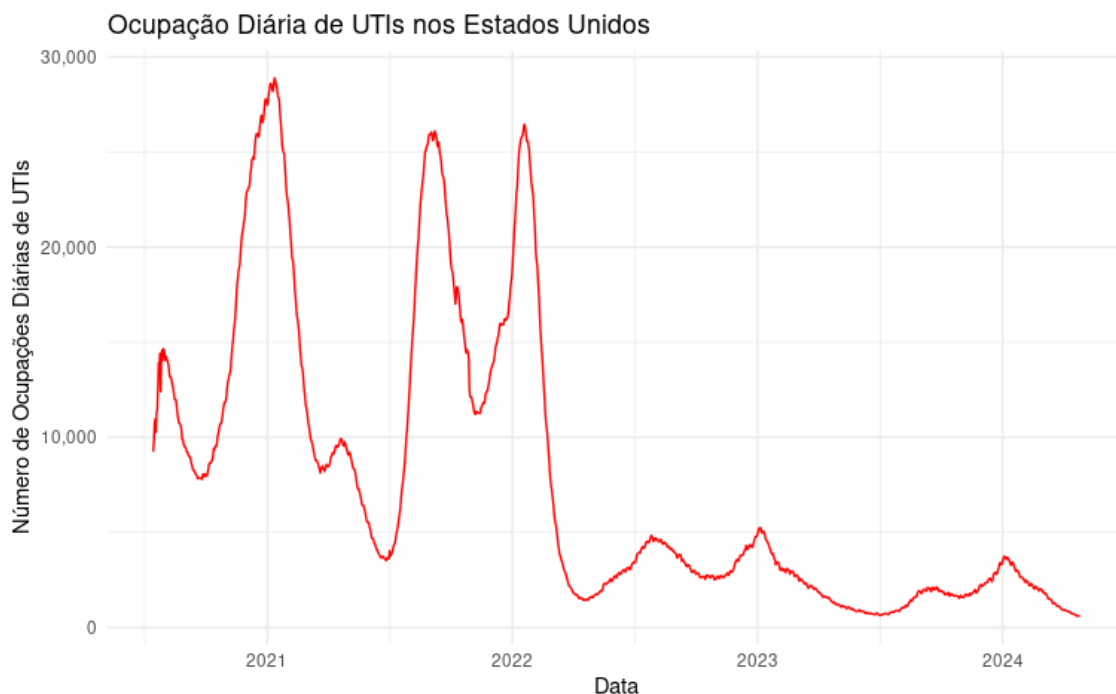
Fonte: Os autores.

A Figura 3 apresenta um ranking dos países com a maior média de admissões semanais nos hospitais. Essa análise nos permite compreender melhor a carga semanal de casos novos admitidos nos hospitais ao redor do mundo.



Fonte: Os autores.

Os Estados Unidos destacaram-se como o país mais afetado e com as maiores médias de hospitalizações durante a pandemia. Para uma visualização mais clara dessa contagem, a Figura 4 exibe um gráfico com o histórico de ocupação de UTIs nos hospitais, enfocando a situação nos Estados Unidos.



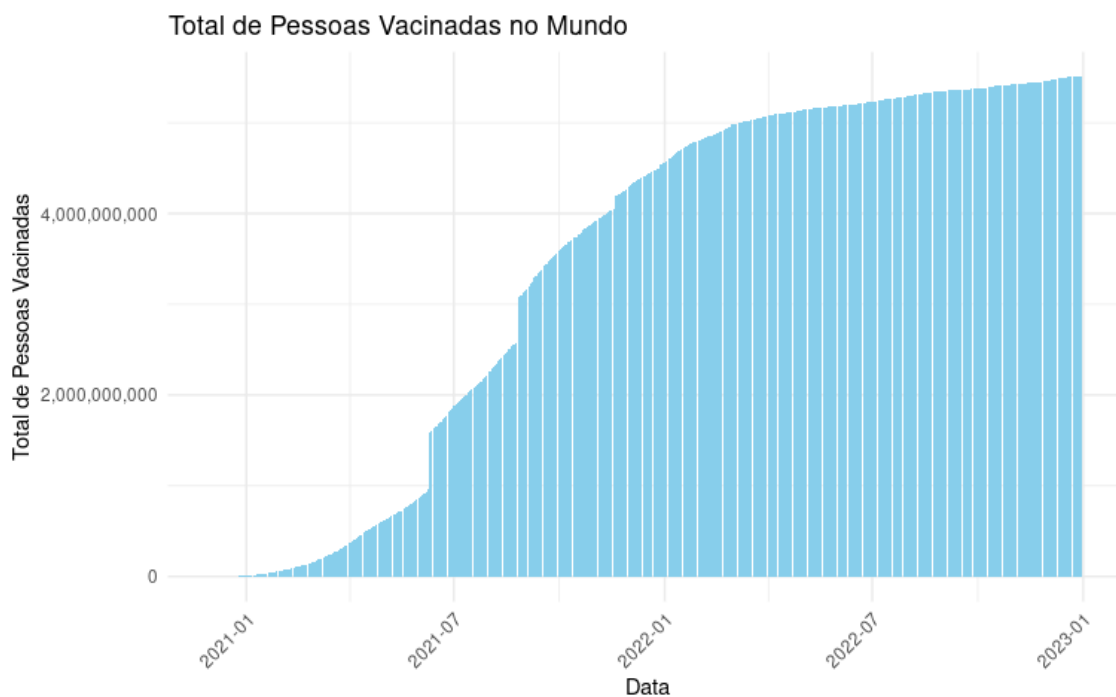
Contagem de Ocupação de UTIs nos Estados Unidos

Fonte: Os autores.

Como é possível observar, a partir do início de 2022, houve uma queda considerável no número de hospitalizações graves, seguida de uma estabilização desse quadro até os dias atuais. Na próxima seção, iremos analisar o impacto da vacinação durante esse mesmo período nos Estados Unidos.

2.0.5 Análise das Vacinações

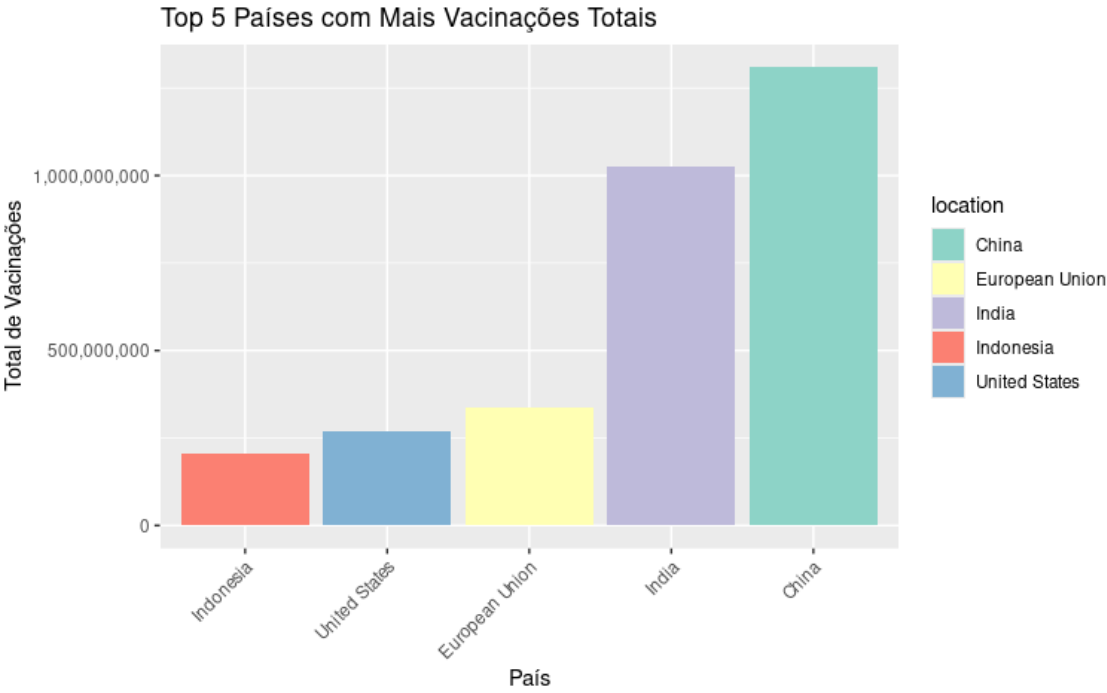
A Figura 1 mostra o total de vacinações administradas em todo o mundo desde o início da pandemia. Esse gráfico oferece uma visão geral do progresso da vacinação em escala global, durante o período de janeiro de 2020 até janeiro de 2023.



Total de Pessoas Vacinadas no Mundo

Fonte: Os autores.

Na Figura 2, apresentamos o ranking dos cinco países que mais vacinaram suas populações. Esse gráfico destaca os países líderes em termos absolutos de doses administradas, refletindo seus esforços para imunizar suas populações contra o vírus, no mesmo período.

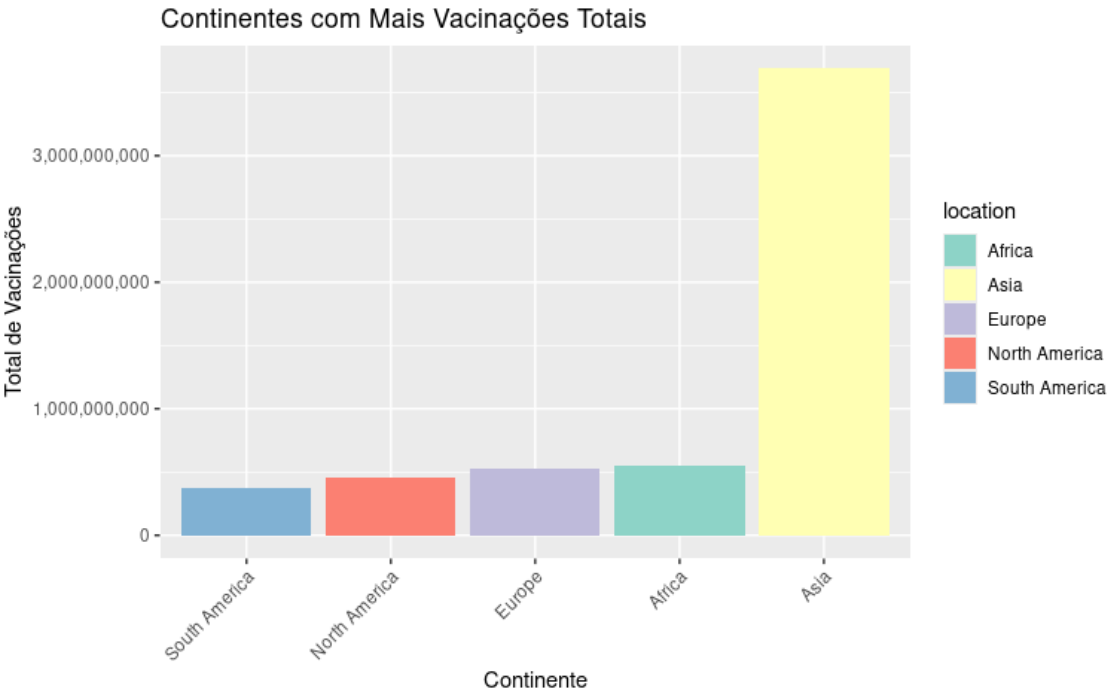


5

Países com Mais Vacinações Totais

Fonte: Os autores.

A Figura 3 exibe o ranking dos cinco continentes com o maior número de pessoas vacinadas. Destaca-se que a Ásia, representando 65% da população mundial, desempenha um papel significativo nesse cenário, influenciando diretamente os números de vacinação global.

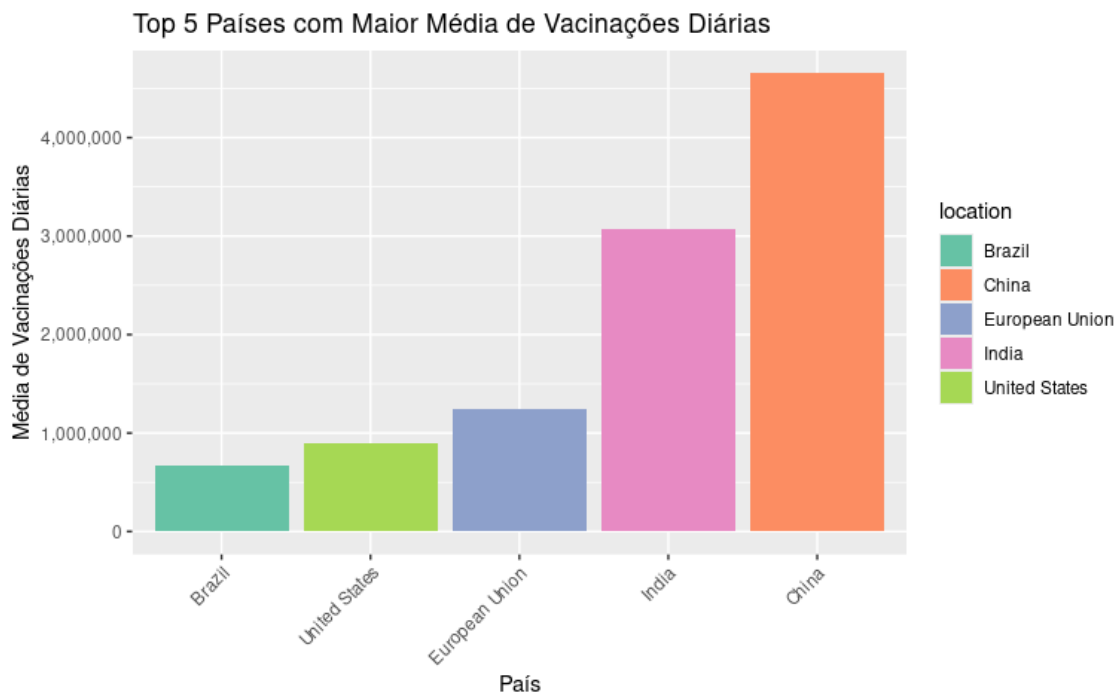


5

Continentes com Mais Vacinações Totais

Fonte: Os autores.

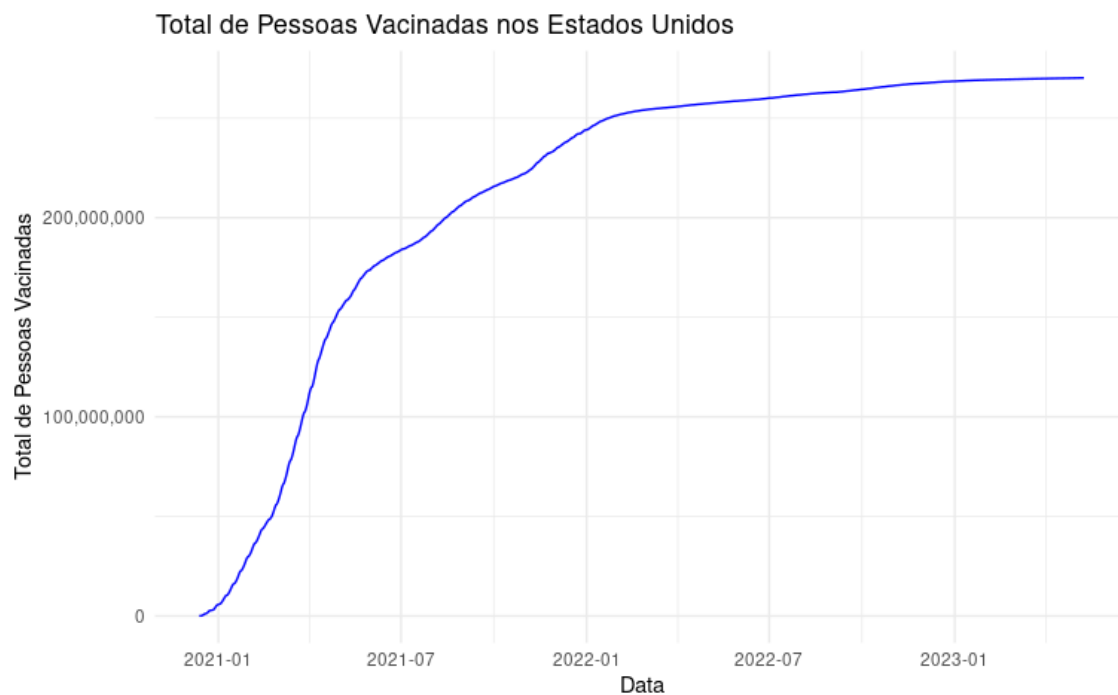
Na Figura 4, analisamos os países com as maiores médias de vacinação diária. Esses países demonstram uma capacidade excepcional de administrar doses de vacinas diariamente, contribuindo significativamente para a imunização de suas populações contra a COVID-19.



Maiores Médias de Vacinação Diária

Fonte: Os autores.

Por fim, na Figura 5, analisamos o histórico do total de pessoas vacinadas apenas nos Estados Unidos, que se destacou anteriormente como país com maior número de hospitalizações.



Histórico de Pessoas Vacinadas nos Estados Unidos

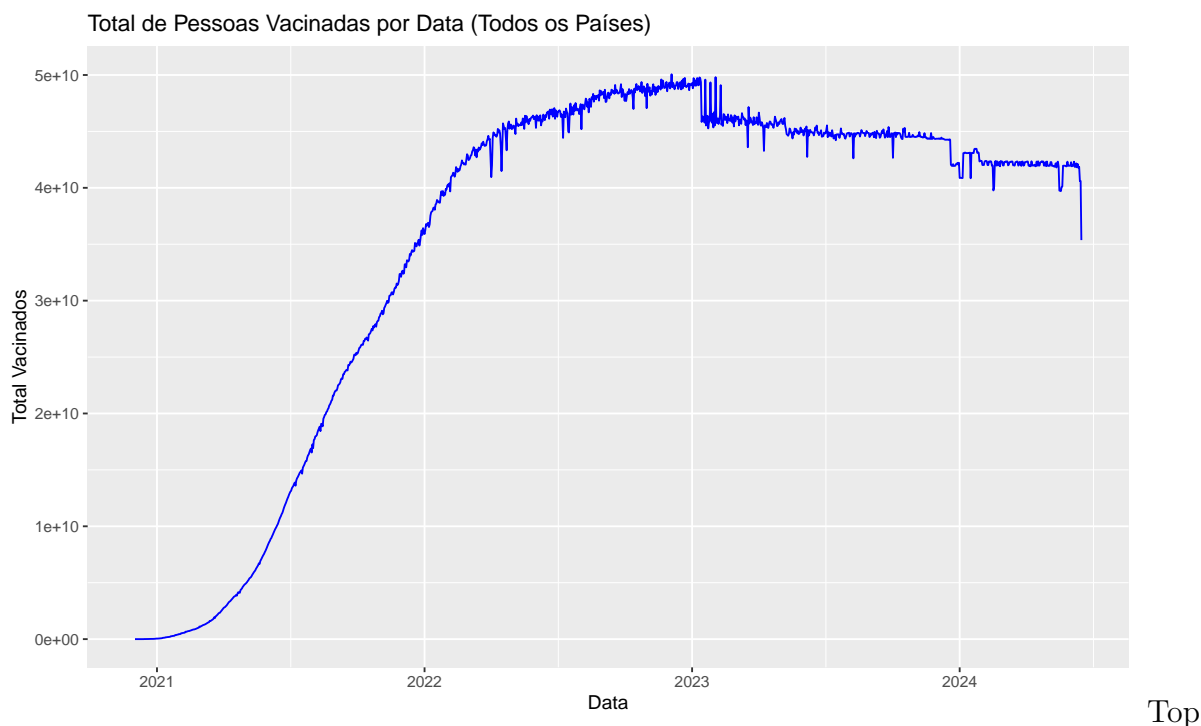
Fonte: Os autores.

Podemos notar uma correlação entre a vacinação em massa no início de 2022 e a subsequente redução no número de internações graves. A estabilização no número de vacinações coincide com uma queda significativa no número de casos graves, sugerindo que a imunização desempenhou um papel importante na redução da gravidade da doença e na pressão sobre os sistemas de saúde.

Este padrão sugere uma relação direta entre a vacinação e a diminuição das hospitalizações por COVID-19, destacando a eficácia das campanhas de vacinação na proteção da população contra casos graves da doença. float

3 Inferência

3.1 Vacinação



10 Países com Maior Média de Ocupação Diária de UTI

Fonte: Os autores.

O gráfico acima, ilustrado com uma linha azul crescente, revela o panorama geral da campanha de vacinação contra a COVID-19, mostrando o total de pessoas vacinadas em todo o mundo ao longo do tempo. A partir dos dados, podemos observar um aumento constante no número de indivíduos imunizados, demonstrando o progresso na luta contra a pandemia.

3.2 Hospitalizações

3.2.1 Teste de Normalidade

O teste de Shapiro-Wilk para normalidade aplicado à ocupação de UTI em casos de COVID-19 ($p\text{-valor} = 0.002$) indica que não podemos rejeitar a hipótese de normalidade ($p\text{-valor} > 0.05$). Isso significa que os dados se distribuem de forma semelhante a uma curva em forma de sino, o que facilita a aplicação de técnicas estatísticas paramétricas.

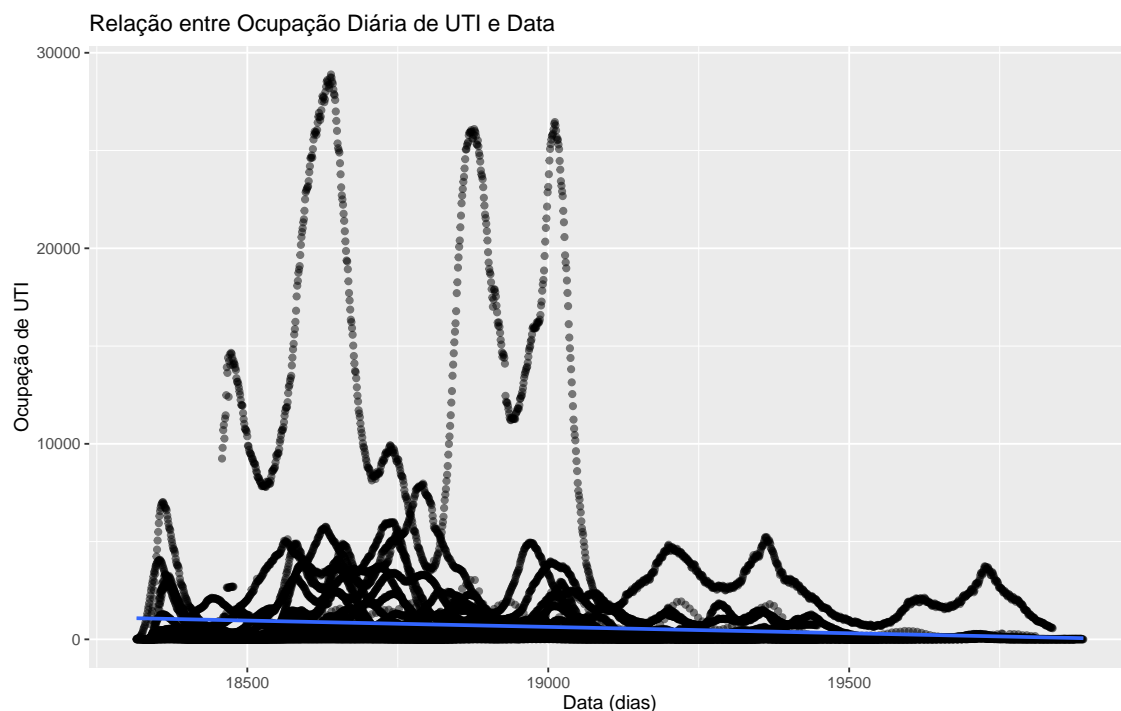
3.2.2 Comparação entre 2022 e 2023

Ao realizar o teste t para médias independentes, constatamos que há uma diferença estatisticamente significativa ($p\text{-valor} < 0.05$) entre as médias de internação por COVID-19 em 2022 e 2023. Isso sugere que, em média, o número de internações em 2023 foi menor do que em 2022. Essa informação pode estar relacionada a diversos fatores, como o avanço da vacinação, a implementação de medidas de controle mais eficazes e a sazonalidade da doença.

3.2.3 Correlação entre Ocupação de UTI e Data

O teste de Correlação de Spearman revelou uma correlação estatisticamente significativa ($p\text{-valor} < 0.05$) e negativa (coeficiente de correlação = -0.34) entre a ocupação de UTI e a data. Essa descoberta indica que, ao longo do tempo, a ocupação de UTI tende a diminuir. Essa tendência pode ser atribuída aos fatores mencionados anteriormente, como o avanço da vacinação e o aprimoramento das medidas de controle da pandemia.

3.2.4 Modelo de Regressão



Top

10 Países com Maior Média de Ocupação Diária de UTI

Fonte: Os autores.

O modelo de regressão linear confirma a relação negativa entre a ocupação de UTI e o tempo, com um coeficiente de regressão beta negativo e estatisticamente significativo. Isso reforça a ideia de que a ocupação de UTI tende a diminuir ao longo do tempo.

A análise estatística dos dados de internações por COVID-19 revelou informações sobre a evolução da pandemia e o impacto das medidas de controle. A diminuição da ocupação de UTI ao longo do tempo, associada à correlação negativa entre essa variável e a data, indica que os esforços para combater a pandemia estão tendo efeito positivo.

4 Probabilidade

4.0.1 Estudo de Probabilidade

A análise da probabilidade de estar vacinado tornou-se essencial ao estudar o avanço da vacinação durante os anos, abrangendo tanto indivíduos que receberam apenas uma dose quanto os completamente vacinados. Esses dados proporcionam insights cruciais sobre a prevalência da vacinação na população. Este estudo revela a importância dessas informações, especialmente quando relacionadas ao número de hospitalizações por COVID-19. É amplamente reconhecido que a vacinação desempenha um papel crucial na redução das hospitalizações, uma vez que há uma relação direta entre a cobertura vacinal e a diminuição da gravidade dos casos. A análise exploratória dos dados revela que à medida que mais pessoas são vacinadas, a probabilidade de uma pessoa estar protegida contra a necessidade de hospitalização aumenta significativamente. Este fenômeno é fundamental para entender o impacto das políticas de vacinação e saúde pública na resposta à pandemia.

4.0.2 Dados sobre a vacinação

A vacinação contra a COVID-19 emergiu como uma resposta crucial para conter a propagação da pandemia globalmente. Este artigo explora o progresso significativo alcançado através da vacinação, destacando o número de pessoas vacinadas e a probabilidade de estar vacinado em relação à população mundial.

Desde o início de 2020 até os dias atuais, observamos um aumento substancial no número de pessoas vacinadas. Em 2020, aproximadamente 7.230.216 indivíduos receberam pelo menos uma dose da vacina, marcando o início da campanha de imunização em escala global. Em contraste, em 2024, o número de pessoas vacinadas atingiu a impressionante marca de 5.631.262.670, refletindo um esforço coordenado e massivo para proteger as populações contra a COVID-19.

A análise da probabilidade de estar vacinado revela insights adicionais sobre a cobertura vacinal em diferentes períodos. Em 2020, a probabilidade era relativamente baixa, refletindo os desafios iniciais de distribuição e acesso às vacinas. No entanto, conforme as campanhas de vacinação foram ampliadas e otimizadas, a probabilidade de uma pessoa estar vacinada aumentou significativamente ao longo dos anos subsequentes. Em 2024, a maioria da população mundial tinha recebido pelo menos uma dose da vacina, ilustrando os esforços contínuos para alcançar a imunidade coletiva contra o vírus.

Além disso, o número de pessoas completamente vacinadas com duas doses também

cresceu exponencialmente ao longo dos anos. Em 2024, mais de 5 bilhões de pessoas haviam completado seu esquema vacinal, o que é crucial para fortalecer a proteção individual e coletiva contra a COVID-19.

A análise feita sobre a probabilidade de a pessoa estar vacinada pode ter sua evolução observada na figura abaixo.



Fonte: Os autores.

5 Conclusão

5.0.1 Desvendando a Relação entre Vacinação e Hospitalizações

Ao longo deste relatório, mergulhamos em uma análise profunda dos dados de vacinação e hospitalização relacionados à COVID-19, desde os primórdios da pandemia em 2020 até os dias atuais. Munidos de informações confiáveis e atualizadas do repositório covid-19-data da Our World in Data (OWID), exploramos as tendências, desigualdades e correlações que moldaram o cenário global da pandemia.

5.0.2 Um Panorama de Hospitalizações

A análise descritiva das hospitalizações revelou disparidades alarmantes. Os Estados Unidos emergiram como o epicentro da crise, com médias de ocupação de UTIs e admissões hospitalares semanais que eclipsaram outros países (Figuras 1 e 2). Essa realidade sombria expõe as falhas e desigualdades nos sistemas de saúde, e serve como um lembrete da gravidade da pandemia.

5.0.3 Vacinação

Em contraste com o panorama das hospitalizações, a análise da vacinação trouxe um raio de esperança. Países como China, Estados Unidos, Índia e Brasil se destacaram em números absolutos de doses administradas (Figura 2), demonstrando um compromisso com a imunização em massa. No entanto, a Figura 3 serve como um alerta: a Ásia, lar da maior parcela da população mundial, ainda tem um longo caminho a percorrer para alcançar a cobertura vacinal ideal.

5.0.4 Vacinação e Redução de Hospitalizações

Um achado emergiu da análise dos dados de vacinação e hospitalização nos Estados Unidos (Figuras 4 e 5). A partir do início de 2022, com o avanço da vacinação em massa, observou-se um declínio notável no número de internações graves por COVID-19. Essa correlação positiva sugere que as campanhas de vacinação desempenharam um papel crucial na mitigação da gravidade da doença e na redução da pressão sobre os sistemas de saúde.

A vacinação é uma ferramenta fundamental para controlar a pandemia de COVID-19. A imunização em larga escala contribui para a redução de casos graves, diminuindo a demanda por leitos de UTI e outros recursos hospitalares, além de salvar vidas.

Diante disso, instamos governos e autoridades de saúde pública a intensificarem seus esforços para ampliar o acesso à vacinação em todo o mundo. Campanhas de conscientização, investimentos em infraestrutura para distribuição de vacinas e combate à desinformação são medidas essenciais para alcançar a imunização global e proteger a saúde pública.

5.0.5 Reconhecendo Limitações

É crucial reconhecer as limitações deste estudo. A análise se baseou em dados agregados por país e continente, sem considerar as disparidades dentro de cada região. Além disso, outros fatores além da vacinação podem ter influenciado as taxas de hospitalização, como medidas de distanciamento social, uso de máscaras e acesso a testes.

Pesquisas futuras podem se aprofundar na análise de dados granulares, considerando fatores socioeconômicos, demográficos e epidemiológicos específicos de cada região. Estudos que explorem a efetividade de diferentes tipos de vacinas e a relação entre a vacinação e a mortalidade por COVID-19 também podem trazer contribuições importantes para o conhecimento científico e para a formulação de políticas públicas mais eficazes no combate à pandemia.

6 Referência bibliográfica

Our World In Data. **About** [Online].
Disponível em: <<https://ourworldindata.org/about>>.
Acesso em: 12 jun. 2024.